DOCUMENT RESUME

ED 185 086 TM 800 080

AUTHOR Araujo, John: Semb, George
TITLE The Effects of Item Order on Student Performance.
PUB DATE [Sep 79]
NOTE 19p.: Paper presented at the Annual Meeting of the American Psychological Association (87th, New York, NY, September 1-5, 1979).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Analysis of Variance: *Course Organization: Higher Education: Lecture Method: Mastery Learning: *Performance Factors: Scheduling: *Scores: *Teaching Methods: Test Construction: *Test Items
IDENTIFIERS *Final Examinations: *Item Position (Tests): Personalized System of Instruction

ABSTRACT

Two experiments used two 2 x 3 x 2 (instructors x test forms x class times) designs to compare three aspects of student performance (mean final examination scores, homogeneity of variance, and retake frequencies) on three test forms. The test items on the three forms were arranged in three ways: (1) forward--items paralleled the presentation of material in class: (2) backward--items were in an opposite order: and (3) random order. The 336 undergraduates were taught using either Contingency Managed Lecture (CML) or the Keller Personalized System of Instruction (PSI). Analyses indicated that: (1) item order alone did not significantly influence student performance in either experiment: (2) test form and class time significantly interacted for CML: (3) none of the item orders affected retake rates in either experiment: (4) analysis of variance of all twelve cells for each experiment revealed heterogeneity only for CML and only by collapsing across the time variable: and (5) the CML experiment also revealed an instructor x test form interaction. It was concluded that there was no evidence that test order disrupted performance, in spite of the fact that there was tendency for random order to produce the best results. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED185086

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Effects of Item Order on Student Performance

John Araujo and George Semb

Department of Human Development

University of Kansas

Lawrence, Kansas 66045

(913) 864 - 4840

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

John Araujo

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper Presented in Division 2 of the American

Psychological Association - 1979

TM800 080

Abstract

Two experiments utilized two, 2 x 3 x 2 (instructors x forms x class times) designs to compare three types of student performances (mean scores, homogeneity of variance, and retake frequencies) on three final examination forms which varied in the way test items were ordered. Students in Experiment I were taught using Contingency Managed Lecture (CML), and students in Experiment II were taught using the Personalized System of Instruction (PSI). Items on the final exam for both experiments were ordered in three ways: (1) forward - test items paralleled the material presented; (2) backward - test items were in an order opposite that of the material presented; and (3) random - test items were randomly ordered. The results indicate that: (1) item order alone does not significantly influence student performances under CML or PSI; (2) form and class time significantly interact for CML; (3) none of the item orders produced significantly different retakes rates for either experiment; (4) variance analyses of all 12 cells for each experiment revealed heterogeneity only for Experiment I, and further analyses revealed that this variability was produced by collapsing across the time variable; and finally, (5) after the CML data underwent a Log X transformation, there was a significant instructor x form interaction and, as before, a significant time x form interaction. Since item orders did not influence any of the three measures, it seems reasonable, therefore, too recommend employing various test-item orders.

The Effects of Item Order on Student Performance

Instructors may assume that the order of questions on a final exam should parallel the order of the material presented in a course (Schmitt & Scheirer, 1977). That is, the first few questions should relate to the concepts earliest in the course, and the test should then proceed to sample a student's repertoire in the same order in which the material was learned. A second assmption is that the same test items can be assembled in various orders to minimize cheating (Spiers & Pihl, 1976). Thus, in spite of different item orders, all students would receive the same questions. However, various item-orders might lead to differential performances.

Memory theory seems to support parallel test construction (Schmitt & Scheirer, 1977; Spiers & Pihl, 1976). Students' performances, if memory theory is predictive, should be better on exams in which the question-order parallels the sequence of course material than on exams with a non-parallel test question-order. However, Hughes, Prytula, and Schnelle (1974), Schmitt and Schierer (1977), and Spiers and Pihl (1976) obtained no significant diffrence in final exam performance whether the order of questions was random or consistent with that of course presentation. When compared with these two types of item sequences, Hughes et al. (1974) and Schmitt and Schierer (1977) also obtained no significant differences when test questions were in reverse order from course presentation.

This study evaluated three final examination orders under two

teaching procedures. In Experiment I, students were taught using Contingency Managed Lecture, (CML, Philips & Semb, 1977) and in Experiment II, students were taught using the Personalized System of Instruction, PSI (Keller, 1968). Three types of performances were evaluated: (1) mean scores, (2) variability, and (3) retake frequencies. Therefore, this study was conducted to systematically replicate and extend previous findings but with well established and researched instructional technologies (Journal of Personalized Instruction, 1977 to present).

Method

Subjects, Setting, and Instructional Staff. Students (N = 336) in eight sections of an introductory child psychology course participated in the study. Two sections (Experiment I) each (N = 42) met at 9:30 a.m. and 11:00 a.m. Tuesday-Thursday (TR), and two sections (Experiment II) each (N = 42) met at 10:30 a.m. and 11:30 a.m. Monday-Wednesday-Friday (MWF). Within each class hour, students were randomly assigned to sections. One instructor and five proctors taught each MWF section, and one instructor and two proctors taught each TR section.

General Course Organization. The course was divided into three major content parts. Each content part consisted of four unit quizzes and a review exam. At the end of the course, students took a comprehensive final exam.

A typical unit consisted of about 30 pages of reading from the course text (LeFrancois, 1977) and 25 short essay questions and several objective questions in the study guide (Semb, 1977).

A unit quiz (worth 10 pts.) consisted of 6-8 one-point, multiple-

choice questions and 1-3 two-point, short-essay questions based on major concepts emphasized in the study guide. The review exam consisted of 24 one-point, multiple-choice questions. The comprehensive final was 60 one-point, multiple-choice questions.

Teaching Procedures, Experiment I. Students in this experiment were taught using Contingency Managed Lecture (Phillips & Semb, 1977). There was one lecture and/or movie prior to each unit quiz, review exam, and the final exam, and lecture attendance was optional. On the class day after every second unit lecture, students took the two unit quizzes which corresponded to those two preceeding lectures and/or movies. Students took their review and final exams the class day following the day on which the review/final lecture was given.

Students could retake all unit quizzes, review exams, and the final exam once. Their highest scores counted toward their grades. If students failed to write a unit quiz or review exam the first time it was administered, the highest possible scores were 7 and 20 for the unit quizzes and review exams respectively. There was no penalty for not writing the final examination the first time it was administered.

Teaching Procedures, Experiment II. Students in this experiment were taught using the Personalized System of Instruction (Keller, 1968). They proceeded at their own pace throughout the course provided they did not fall below a uniform rate line that distributed the work evenly throughout the semester. Students whose rates fell below this line entered daily testing. Daily testing meant that the student was required to test every class day until they caught up with the uniform rate line.

Student progress required a 100% mastery (10 pts.) on each quiz. Students, who received less than 10 but more than 7 points, were able to correct, in writing, their incorrect answers. However, students who received less than 8 points had to retake alternate forms of the same unit's quiz until they wrote a 10 point quiz.

The review exam could be retaken as many times as the student wanted, provided the student also complied with the uniform rate line and daily testing procedures. However, before each odd attempt (third, fifth, seventh, and so on), the student had to redo the immediately preceeding four units. The best review exam score counted towards the student's grade.

The final exam, which was taken after the third review exam and which marked the end of the course, also could be taken twice, and the best score counted towards the student's grade.

Final Test Construction, Experiments I and II. There were three forms of the final exam, and each form contained 5 questions from each of the 12 units. These 5 questions were randomly selected from a pool of 30 questions per unit. The forms differed in the sequence in which test items appeared: (1) forward - The questions appeared on the final in the same order in which they were presented throughout the course. Thus, Unit 1 questions appeared first and Unit 12 questions were last; (2) backward - The questions appeared in the opposite order to that in which they were presented, i.e., Unit 12 questions appeared first, and Unit 1 questions were at the end of the exam; and (3) random - The questions were randomly ordered.

Experimental Design and Assignment of Students to Test Forms. A 2 x 3 x 2 (2 instructors x 3 forms x 2 class times) factorial design was used to evaluate the effects of the three test question-orders under each teaching procedure, CML and PSI. During each class time, the final exam forms were randomly assigned to each student, and each student completed only one form. One hundred-forty seven students (49 = forward, 46 = backward, 52 = random) took the final exam in Experiment I, and 126 students (31 = forward, 40 = backward, 47 = random) took the final exam in Experiment II. Thus, the final N for analysis was 273 students. While the final exam could be taken twice, an order-form was assigned to only the first attempt. All retakes were the forward form.

Pairs of Instructors were crossed within experiments but nested between experiments. Forms and times were fully crossed. While PSI class times met at 10:30 a.m. and 11:30 a.m., times 1 and 2 respectively, and CML class times met at 9:30 a.m. and 11:00 a.m., also times 1 and 2 respectively, it was felt that these hour and half-hour differences respectively, would still allow the time variable to be considered fully crossed.

Dependent Variable. The dependent variable was student performance on one of the three final exam forms. Students' responses were entered on a standardized response sheet, and these sheets were scored with a template.

Results

Experiment I, CML Instruction. Prior to all analyses and by cells, all departures from normality at the tails (extreme cases) were

evaluated for extremity by Dixon's test (Dixon, 1953; Dixon & Massey, 1957, pp. 275 ff.), and extreme scores were Winsorized (Miller, 1966, p. 214; Wike, 1971). For Experiment I, this analysis revealed that there were two extreme scores. An 18 (instructor 1 at 9:30 a.m.) was scored on a forward-form, and a 25 (instructor 2 also at 9:30 a.m.) was scored on a random-form. Both the extreme scores were Winsorized (replaced with the closest non-extreme score) with a 41.

The mean score data gathered during Experiment I are depicted in Table 1. The mean performance was 47.53 correct answers from a possible of 60, and the scores ranged from 20 to 60 after the extreme case analyses were conducted.

Insert Table 1 about here

The results of the three-way, fixed-factor analysis of variance for CML Instruction are presented in Table 2.

Insert Table 2 about here

This analysis only revealed a significant time x form interaction, F (2, 135) = 3.83, $p < .05$. There were no significant instructor, time, or form main effects, and no other significant interactions. Both time and form simple main effects were analyzed. These analyses revealed that

none of the simple main effects were significant.

Tests of homogeneity of variance (Bartlett, 1937) were conducted for the entire 2 x 3 x 2 matrix as well as for collapsing across each factor. Heterogeneity of variance was significant for the entire matrix, F (11, 12385) = 2.63, $p < .05$, as well as for collapsing across the time factor, F (5, 25384) = 2.73, $p < .05$.

Retake frequencies for the three forms were calculated and analyzed with a simple, fixed-factor analysis of variance. The forward, backward, and random forms produced 0.41, 0.37, and 0.33 mean retakes respectively. The analysis of variance was not significant, F (2, 9) = 0.31, $p > .05$.

Experiment II, PSI Instruction. As in Experiment I, Dixon's (1953) test was applied to suspected extreme scores. However, these analyses revealed no extreme scores.

The data gathered during Experiment II are presented in Table 2. The mean performance was 49.73 correct and scores ranged from 28 to 59. The results of the three-way, fixed-factor analysis of variance for PSI Instruction are presented in Table 3.

Insert Table 3 about here

Except that the instructor x form interaction approached significance, F (1, 14) = 3.18, $p = .077$, there were no significant findings for this analysis. Furthermore, there was homogeneity of variance amongst the 12

cells as well as for collapsing across each factor.

Since students had the opportunity to retake the final exam once, retake frequencies were analyzed for each of the three forms. Backward, forward, and random forms produced 0.42, 0.35, and 0.25 mean retakes respectively. The analysis of variance was not significant, F (2, 9) = 3.07, $p > .05$.

Discussion

CML Instruction. Under CML instruction, the three final exam forms alone did not produce significantly different mean performances. However, it is interesting to note that the forward form produced the best performance for only one out of four classes or in only one out of 12 cells in this experiment. This was for instructor 1 at 11:00 a.m. These data do not support the assumption that question-orders that do not parallel the order of the material learned may interrupt the acquired behavioral chain. Nor do they not support memory theory. These findings are consisted with previous research that has also shown no differences due to the order of items (Hughes, et al. 1974; Schmitt & Scheirer, 1977; and Spiers & Pihl, 1976).

What can be recommended for instructors and students involved in contingency managed lecture teaching? First, final exam forms alone do not significantly influence mean performances. However, if these performances are compared at different times or with various instructors, then they can be altered. Second, forms alone do not affect homogeneity of variance. However, several factors combined might, and therefore, distributions should be constructed for students who can be classified

according to these factors. Finally, various forms do not produce significantly different retake rates. Therefore, the order of multiple choice items alone on final exams for CML classes appears to have an insignificant effect on student performances.

PSI Instruction. Under PSI instruction, the three-way analysis of variance revealed no significant findings. However, inspite of the non-significant findings, the random form produced the best results for three out of four instructors. Furthermore, when the mean performances for each order were collapsed across instructors for each hour, students performed best with the random order and worst with the forward order, with the backward order falling in between. These data further reduce support for the assumption that test item orders should parallel the order of the material learned. However, why the random form should consistently produce better performances is not clear. Such results seem to suggest that there is no logical manner in which tests should be constructed.

The mastery component of PSI would seem to produce a ceiling effect in terms of the material learned so that regardless of how the items were order, performances would be equal. Perhaps students rarely answer questions sequentially. Occaisionally, or frequently, they skip difficult questions to return to them later. The new effect of this strategy may be to randomize the items. Future research might attempt to control for this type of student strategy by requiring students to answer questions sequentially. In other words, testing would progress contingent upon answering questions as they are presented.

The recommendation for PSI instructors and students is even more clear than those for CML. PSI appears to be a robust instructional technique unaffected in terms of mean scores, variability, or retake frequencies by either final exam forms alone or in combination with other factors.

Conclusions.

Originally, it was hypothesized that test item orders which departed from the order of material learned would disrupt three types of performances. These were: (1) means, (2) variability, and (3) retake frequencies. Orderly studying combined with a sequence of 12 unit quizzes might have produced a performance chain which could be disrupted by testing in an order different, presumably, from that learned. However, data gathered in two experiments found no evidence that test orders can disrupt mean performances in spite of the fact that there was a tendency for a random order to produce the best mean performances. The reasons for this finding and a suggestion for future research were discussed.

If item orders would not disrupt mean performances, would they disrupt variablility and/or produce differential retake frequencies? Apparently not. Variability was not influence by various item-orders, nor were retake frequencies.

References

Bartlett, M. S. Properties of sufficiency and statistical tests. Proc. Ray. Soc. (London), Ser. A, 1937, 160, 268-282.

Box, G. E. P. Non-normality and tests on variances. Biometrika, 1953, 40, 318-335.

Dixon, W. J., & Massey, F. J., Jr. Introduction to Statistical Analyses. New York: McGraw-Hill, 1951.

Dixon, W. J. Processing data for outliers. Biometrics, 1953, 9, 74-89.

Hughes, G. E., Prytula, R. E., & Schnelle, J. F. Relationship between a forward learning set and three orders of multiple-choice examinations. Psychological Reports, 1974, 34, 727-733.

Keller, F. S. "Good-bye, teacher . . ." Journal of Applied Behavior Analysis, 1968, 1, 79-89.

LeFrancois, F. R. Of Children. Belmont, CA: Wadsworth, 1977.

Miller, R. G., Jr. Simultaneous Statistical Inference. New York: McGraw-Hill, 1966.

Philips, T. W., & Semb, G. Using peer grading to personalize a contingency-managed lecture course. Journal of Personalized Instruction, 1977, 2, 150-155.

Scheffe, H. A method for judging all contrasts in the analysis of variance. Biometrika, 1953, 40, 87-104.

Schmitt, J. C., & Scheirer, C. J. The effect of item order on objective tests. Teaching of Psychology, 1977, 4, 144-145.

Semb, G. Study Guide for Of Children. Belmont, CA: Wadsworth, 1977.

Spiers, P. A., & Pihl, R. O. The effect of study habits, personality

and order of presentation on success in an open-book objective examination. Teaching of Psychology, 1976, 3, 33-34.

Wike, E. L. Data Analysis: A Statistical Primer for Psychology Students. Chicago: Aldine-Atherton, 1971.

Table 1

Mean Scores for Each of the Three-Way Classification Experiments

Results for CML Instruction

Question-Order Forms at 9:30, Time 1

		Forward	Backward	Random	Row Means	Instruction Mean
Instructors	1	45.31	50.09	44.23	46.35	
		±3.77	±7.41	±12.08		
	2	43.09	49.33	47.85	46.89	
		±1.35	±5.14	±4.36		
Column Means		44.29	49.70	46.04	46.62	

Question-Order Forms at 11:00, Time 2

		Forward	Backward	Random	Row Means	
Instructors	1	52.09	48.67	49.42	50.00	
		±5.58	±7.82	±7.12		
	2	45.36	44.18	50.93	47.03	
		±8.75	±6.59	±7.92		
Column Means		48.32	46.52	50.23	48.43	47.53

Results for PSI Instruction

Question-Order Forms at 10:30, Time 1

		Forward	Backward	Random	Row Means	Instruction Mean
Instructors	3	48.27	49.92	50.00	49.44	
		±6.90	±5.85	±5.58		
	4	51.83	50.88	53.55	52.28	
		±4.79	±5.84	±3.50		
Column Means		49.53	50.30	51.63	50.61	

Table 1 (continued)

Question-Order Forms at 11:30, Time 2

		Forward	Backward	Random	Row Means	Instruction Mean
Instructors	3	51.00 ±6.34	47.75 ±6.04	50.20 ±5.77	49.79	
	4	46.08 ±9.25	49.25 ±6.90	49.31 ±6.28	48.24	
Column Means		48.32	48.65	49.70	48.91	49.73

Note. Cell format is: mean
± S.D.

Table 2

Summary Table for the Three-Way ANOVA For CML Instruction

Source[a]	Sum of Squares	Degrees of Freedom	Mean Square	F
I	83.25	1	83.25	1.39
T	116.88	1	116.88	1.96
F	85.73	2	42.86	0.72
I X T	108.46	1	108.46	1.82
I X F	335.32	2	167.66	2.81
T X F	457.05	2	228.52	3.83*
I X T X F	9.53	2	4.77	0.08
Error	8063.53	135	59.73	
Total	9259.75	146		

[a]I = Instructor, T = Time, F = Form.

*$p < .05$.

Table 3

Summary Table for the Three-Way ANOVA For PSI Instruction

Source[a]	Sum of Squares	Degrees of Freedom	Mean Square	F
I	11.72986	1	11.72986	0.29
T	97.95081	1	97.95081	2.45
F	55.60956	2	27.80478	0.70
I X T	127.33618	1	127.33618	3.18
I X F	24.47021	2	12.23511	0.31
T X F	1.37646	2	0.68823	0.02
I X T X F	95.48877	2	47.74438	1.19
Error	4558.57001	114	39.98746	
Total	4972.53186	125		

[a]I=Instructor, T=Time, and F=Form.